

Rec'd PCT/PTO 13 FEB 2004

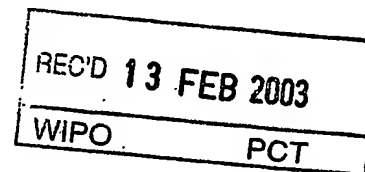
10/50/2788  
PCT/EP 03 / 00003



Europäisches  
Patentamt

European  
Patent Office

Office européen  
des brevets



Bescheinigung

Certificate

Attestation

Die angehefteten Unterla-  
gen stimmen mit der  
ursprünglich eingereichten  
Fassung der auf dem näch-  
sten Blatt bezeichneten  
europäischen Patentanmel-  
dung überein.

The attached documents  
are exact copies of the  
European patent application  
described on the following  
page, as originally filed.

Les documents fixés à  
cette attestation sont  
conformes à la version  
initialement déposée de  
la demande de brevet  
européen spécifiée à la  
page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

02001256.3

**PRIORITY  
DOCUMENT**  
SUBMITTED OR TRANSMITTED IN  
COMPLIANCE WITH RULE 17.1(a) OR (b)

Der Präsident des Europäischen Patentamts;  
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets  
p.o.

R C van Dijk



Europäisches  
Patentamt

European  
Patent Office

Office européen  
des brevets

Anmeldung Nr:  
Application no.: 02001256.3  
Demande no:

Anmeldetag:  
Date of filing: 17.01.02  
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

SIEMENS AKTIENGESELLSCHAFT  
Wittelsbacherplatz 2  
80333 München  
ALLEMAGNE

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:  
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.  
If no title is shown please refer to the description.  
Si aucun titre n'est indiqué se référer à la description.)

Betriebsverfahren eines automatischen Spracherkenners zur sprecherunabhängigen  
Spracherkennung von Worten aus verschiedenen Sprachen und automatischer  
Spracherkenner

In Anspruch genommene Priorität(en) / Priority(ies) claimed /Priorité(s)  
revendiquée(s)  
Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/  
Classification internationale des brevets:

G10L/

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of  
filing/Etats contractants désignées lors du dépôt:

AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE TR

17. Jan. 2002

## Beschreibung

Betriebsverfahren eines automatischen Spracherkenners zur sprecherunabhängigen Spracherkennung von Worten aus verschiedenen Sprachen und automatischer Spracherkenner

Die Erfindung betrifft ein Betriebsverfahren eines automatischen Spracherkenners zur sprecherunabhängigen Spracherkennung von Worten aus verschiedenen Sprachen gemäß Patentanspruch 1 und einen entsprechenden automatischen Spracherkenner gemäß Patentanspruch 6.

Für die phonembasierte Spracherkennung ist ein Spracherkennungs-Vokabular erforderlich, das die phonetischen Beschreibungen aller zu erkennender Wörter umfaßt. Dies ist eine Grundvoraussetzung für die phonembasierte Spracherkennung. Wörter werden hierbei durch Phonemfolgen oder -ketten im Vokabular repräsentiert. Während eines Spracherkennungs-Vorgangs wird eine Suche nach dem besten Pfad durch die Phonemfolgen im Vokabular durchgeführt. Diese Suche kann beispielsweise mit dem sogenannten Viterbi-Algorithmus erfolgen. Bei kontinuierlicher Spracherkennung können zudem die Wahrscheinlichkeiten für Übergänge zwischen Wörtern modelliert und in den Viterbi-Algorithmus einbezogen werden.

Die phonetischen Umschriften für die zu erkennenden Wörter sind die Basis der phonembasierten Spracherkennung. Daher stellt sich zu Beginn des Einsatzes eines phonembasierten Spracherkenners immer die Frage, wie derartige phonetische Umschriften gewonnen werden können. Unter phonetischen Umschriften werden hier die phonetischen Beschreibungen der Wörter aus einem Zielvokabular verstanden. Insbesondere stellt sich diese Frage bei Wörtern, die dem Spracherkenner nicht bekannt sind.

Bekannt sind Mobil- oder Schnurlostelefone, die eine sprecherabhängige Namenswahl ermöglichen. Ein Benutzer eines der-

artigen Telefons muß hierzu die im elektronischen Telefonbuch des Telefons enthaltenen Einträge trainieren, um diese später zur Namenswahl per Sprache nutzen zu können. Allerdings kann in der Regel kein anderer Benutzer dieses Feature nutzen, da  
5 die sprecherabhängige Namenswahl nur für eine Person geeignet ist, nämlich für diejenige, welche die Sprachwahl trainiert hat. Um dieses Problem zu umgehen, können die Einträge im elektronischen Telefonbuch in phonetische Umschriften umgewandelt werden.

10

Zum Ermitteln der phonetischen Umschrift aus einem geschriebenen Wort, beispielsweise einem Telefonbucheintrag, sind unterschiedliche Ansätze bekannt. Es sei hier beispielsweise auf die sogenannten Diktiersysteme, die im allgemeinen auf  
15 einem PC zur Ausführung kommen, verwiesen. Bei derartigen Diktiersystemen ist im Normalfall ein Lexikon von typischerweise mehreren 10000 Wörtern mit den Zuordnungen von Buchstabenfolgen zu Phonemfolgen hinterlegt. Da ein solches Lexikon allerdings einen sehr hohen Speicherplatzbedarf aufweist, ist  
20 es für mobile Endgeräte wie beispielsweise Mobil- oder Schnurlostelefone nicht praktikabel.

Bekannt sind auch Systeme, bei denen die Umsetzung eines Wortes in dessen phonetische Umschrift regelbasiert oder durch  
25 speziell trainierte neuronale Netze erfolgt. Diese Verfahren besitzen wie das Lexikon den Nachteil, daß festgelegt werden muß, in welcher Sprache die Phonemfolge realisiert werden soll. Allerdings können insbesondere in elektronischen Telefonbüchern Namen aus verschiedenen Sprachen vorhanden sein.  
30 Eine Umsetzung wäre dann mit dem oben beschriebenen Verfahren nicht oder nur unvollständig möglich.

Daher wurden sogenannte multilinguale Systeme zur Phonemkettenermittlung und Spracherkennung entworfen. Diese Systeme  
35 erlauben die Erzeugung von Phonemketten aus verschiedenen Sprachen.

Schließlich existiert noch eine andere Lösung: Ein Benutzer spricht die Worte in ein Spracherkennungssystem ein, das daraus automatisch Phonemfolgen generiert. Bei großen Wortschätzen, aber auch schon bei einigen dutzend Wörtern, wie beispielsweise bei einem elektronischen Telefonbuch mit 80 Einträgen, ist dies für den Benutzer nicht mehr akzeptabel.

Aufgabe der vorliegenden Erfindung ist es daher, ein Betriebsverfahrens eines automatischen Spracherkenners zur sprecherunabhängigen Spracherkennung von Worten aus verschiedenen Sprachen sowie einen entsprechenden automatischen Spracherkenner vorzuschlagen, welche einfach zu implementieren sind, sich insbesondere zum Einsatz in mobilen Endgeräten eignen, und kostengünstig zu realisieren sind.

Diese Aufgabe wird durch ein Betriebsverfahren mit den Merkmalen des Patentanspruchs 1 und durch einen automatischen Spracherkenner mit den Merkmalen des Patentanspruchs 6 gelöst.

Der Erfindung liegt im wesentlichen die Idee zugrunde, phonetische Umschriften von Wörtern jeweils für N-verschiedene Sprachen zu ermitteln, diese anschließend nachzuverarbeiten und einer phonembasierten einsprachigen Spracherkennung zuzuführen. Diese Vorgehensweise beruht im wesentlichen auf der Erkenntnis, daß ein Benutzer der Spracherkennung normalerweise in seiner Muttersprache spricht. Auch fremdsprachige Wörter, beispielsweise Namen, spricht er normalerweise mit einer "Muttersprachenfärbung", also einem Akzent aus, die bzw. der durch einen sogenannten Muttersprachen-Spracherkenner grob modelliert werden kann. Das Betriebsverfahren geht daher von einer als Muttersprache definierten Sprache aus.

Jede Sprache läßt sich nun mit unterschiedlichen, der jeweiligen Sprache eigenen Phonemen beschreiben. Bekanntermaßen ähneln sich jedoch viele Phoneme verschiedener Sprachen. Ein Beispiel hierfür ist das "p" im englischen und deutschen.

Diese Tatsache wird bei der multilingualen Spracherkennung ausgenutzt. Für ein Ensemble von Sprachen wird hier ein einziges Hidden-Markov-Modell erstellt, mit dem simultan mehrere Sprachen erkannt werden können. Allerdings führt dies zu einem sehr großen Hidden-Markov-Modell, das eine niedrigere Erkennungsrate als ein einsprachiges Hidden-Markov-Modell besitzt. Zudem muß bei einer Erweiterung des Ensembles von Sprachen um beispielsweise eine weitere Sprache ein neues Hidden-Markov-Modell erstellt werden, was sehr aufwendig ist. Dies wird mit der Erfindung vermieden.

Erfindungsgemäß werden in einem ersten Schritt der Eingabephase zur Erstellung eines Spracherkenner-Vokabulars eines Betriebsverfahrens eines automatischen Spracherkenners zur sprecherunabhängigen Spracherkennung von Worten aus verschiedenen Sprachen, insbesondere von Erkennung von Namen aus verschiedenen Sprachen, die phonetischen Umschriften von Wörtern jeweils für N-verschiedene Sprachen ermittelt, um pro Wort N-erste Phonemfolgen entsprechend N-ersten Aussprachevarianten zu erhalten. In einem zweiten Schritt werden die Ähnlichkeiten zwischen den Sprachen ausgenutzt. Hierzu wird eine Abbildung der Phoneme jeder Sprache auf den jeweiligen Phonemsatz der Muttersprache implementiert. Ferner wird in einem dritten Schritt die implementierte Abbildung auf die im ersten Schritt ermittelten N-ersten Phonemfolgen für jedes Wort angewandt. Dadurch werden pro Wort N-zweite Phonemfolgen entsprechend N-zweiten Aussprachevarianten erhalten. Mit dem Muttersprachen-Spracherkenner kann dann bereits eine Anzahl N-verschiedener Sprachen nach Erstellen eines Spracherkenner-Vokabulars mit den im vorhergehenden Schritt erhaltenen N-zweiten Phonemfolgen pro Wort für den Muttersprachen-Spracherkenner erkannt werden.

Die Erfindung hat im wesentlichen die folgenden Vorteile:  
Während ein Look-up-Verfahren in einem Lexikon bei mobilen Endgeräten wegen des großen Speicherplatzbedarfs scheitert und bei der multilingualen Spracherkennung, die für einen

Satz von Sprachen optimiert wurde, für jede neue Sprache neue Hidden-Markov-Modelle erstellt und optimiert werden müssen, wird durch die Grapheme/Phoneme-Konversion in mehrere Sprachen gemäß der Erfindung ein multilinguales System geschaffen, das mit relativ einfachen Mitteln zu implementieren ist, sich daher vor allem zum Einsatz in mobilen Endgeräten eignet und nicht zuletzt kostengünstig zu realisieren ist. Für die Erfindung ist neben der Graphem-zu-Phonem-Umwandlung im wesentlichen nur noch ein Mapping, d. h. ein Abbilden zwischen den einzelnen Sprachen - wie oben erläutert - erforderlich. Die Phonemfolgen-Ermittlung und das anschließende Mapping bzw. Abbilden laufen normalerweise "offline" auf einem Gerät ab, beispielsweise einem Mobiltelefon, einem Personal Digital Assistant oder Personal Computer mit entsprechender Software, und sind daher zeitunkritisch. Die hierfür benötigten Ressourcen können in einem langsamen externen Speicher untergebracht werden.

Da das mit dem oben beschriebenen Verfahren erstellte Spracherkenner-Vokabular jedoch für jedes Wort N-Aussprachevarianten umfaßt, ist der Suchaufwand bei der Spracherkennung groß. Um ihn zu verringern, kann ein weiterer Schritt in das Verfahren eingeführt werden, der noch vor dem Erstellen des Spracherkenner-Vokabulars und nach dem Erzeugen der N-zweiten Phonemfolgen pro Wort ausgeführt wird. In diesem Schritt werden die N-zweiten Phonemfolgen entsprechend den N-zweiten Aussprachevarianten jedes Wort bearbeitet, indem jede zweite Phonemfolge mittels geeigneter Distanzen, insbesondere der Levenshtein-Distanz, analysiert und klassifiziert wird, und die N-zweiten Phonemfolgen jedes Wortes auf wenige, vorzugsweise zwei bis drei, Phonemfolgen reduziert werden, insbesondere indem die Aussprachevarianten weggelassen werden, die der Aussprachevariante der Muttersprache am wenigsten ähnlich sind. Vereinfacht ausgedrückt werden durch diese Reduzierung die weniger wichtigen Aussprachevarianten weggelassen, wodurch sich der Suchaufwand bei der Spracherkennung verringert.

Eine weitere Aufwandsreduktion läßt sich erreichen, indem vor dem ersten Schritt eine Sprachidentifikation und -reduktion vorgenommen wird. Im Rahmen dieser Sprachidentifikation wird  
5 für jedes zu erkennende Wort die Wahrscheinlichkeit der Zugehörigkeit zu jeder der N-verschiedenen Sprachen bestimmt. Anhand des Ergebnisses dieser Sprachidentifikation wird die Anzahl der im ersten Verfahrensschritt zu verarbeitenden Sprachen, vorzugsweise auf zwei bis drei verschiedene Sprachen,  
10 reduziert. Diese Sprachreduktion erfolgt vorzugsweise, indem die Sprachen mit der geringsten Wahrscheinlichkeit nicht weiterverarbeitet werden. Für ein bestimmtes Wort kann das Ergebnis der Sprachidentifikation beispielsweise wie folgt lauten: "Deutsch 55%, UK-Englisch 16%, US-Englisch 14%, Schwedisch 3%, ...". Bereits nach diesem Ergebnis kann auf drei  
15 verschiedene Sprachen reduziert werden, indem Schwedisch weggelassen, d. h. nicht weiterverarbeitet wird.

Das Ermitteln der phonetischen Umschriften im ersten Verfahrensschritt erfolgt vorzugsweise durch mindestens ein neuronales Netz. Neuronale Netze haben sich zum Ermitteln phonetischer Umschriften aus geschriebenen Worten bewährt, da sie gute Ergebnisse hinsichtlich der Genauigkeit und vor allem Verarbeitungsgeschwindigkeit liefern sowie einfach, insbesondere  
20 in Software implementierbar sind.

Als Muttersprachen-Spracherkenner kann insbesondere ein Hidden-Markov-Modell zum Einsatz kommen, das für die als Muttersprache definierte Sprache erstellt worden ist.  
30

Die Erfindung betrifft ferner einen Spracherkenner zur sprecherunabhängigen Spracherkennung von Worten aus verschiedenen Sprachen, insbesondere zur Erkennung von Namen aus verschiedenen Sprachen. Hierbei ist eine der verschiedenen Sprachen  
35 als Muttersprache definiert. Der Spracherkenner umfaßt  
- einen Muttersprachen-Spracherkenner,



- ein erstes Verarbeitungsmodul zum Ermitteln der phonetischen Umschriften von Wörtern jeweils für N-verschiedene Sprachen, um pro Wort N-erste Phonemfolgen entsprechend N-ersten Aussprachevarianten zu erhalten,
- 5 - ein zweites Verarbeitungsmodul zum Implementieren einer Abbildung der Phonem jeder Sprache auf dem jeweiligen Phonem-satz der Muttersprache,
- ein drittes Verarbeitungsmodul zum Anwenden der mit dem zweiten Verarbeitungsmodul implementierten Abbildung auf die
- 10 mit dem ersten Verarbeitungsmodul ermittelten N-ersten Phonemfolgen für jedes Wort, wodurch pro Wort N-zweite Phonemfolgen entsprechend N-zweiten Aussprachevarianten erhalten werden, die mit dem Muttersprachen-Spracherkenner erkannt werden können und
- 15 - ein viertes Verarbeitungsmodul zum Erstellen eines Spracherkenner-Vokabulars mit den durch das dritte Verarbeitungsmodul erhaltenen N-zweiten Phonemfolgen pro Wort für den Muttersprachen-Spracherkenner.
- 20 In einer bevorzugten Ausführungsform umfaßt der automatische Spracherkenner ein fünftes Verarbeitungsmodul zum Bearbeiten der N-zweiten Phonemfolgen entsprechend den N-zweiten Aussprachevarianten jedes Wortes. Das fünfte Verarbeitungsmodul ist derart ausgebildet, daß jede zweite Phonemfolge mittels
- 25 geeigneter Distanzen, insbesondere der Levenshtein-Distanz, analysiert und klassifiziert wird, und die N-zweiten Phonemfolgen jedes Wortes auf wenige, vorzugsweise zwei bis drei, Phonemfolgen reduziert werden.
- 30 Ferner kann der automatische Spracherkenner einen Sprachidentifikator und einen Sprachreduzierer umfassen. Der Sprachidentifikator ist vor das erste Verarbeitungsmodul geschaltet und bestimmt für jedes zu erkennende Wort die Wahrscheinlichkeit der Zugehörigkeit zu jeder der N-verschiedenen Sprachen.
- 35 Der Sprachreduzierer reduziert die Anzahl der vom ersten Verarbeitungsmodul zu verarbeitenden Sprachen, vorzugsweise auf zwei bis drei verschiedene Sprachen, indem die Sprachen mit

der geringsten Wahrscheinlichkeit nicht weiterverarbeitet werden. Sprachidentifikator und Sprachreduzierer verringern sowohl den Verarbeitungsaufwand des automatischen Spracherkenners sowohl in der Eingabephase als auch in der Erkennungsphase beträchtlich.

Vorzugsweise weist das erste Verarbeitungsmodul mindestens ein neuronales Netz zum Ermitteln der phonetischen Umschriften auf.

10

Schließlich weist der Muttersprachen-Spracherkenner in einer bevorzugten Ausführungsform ein Hidden-Markov-Modell auf, das für die als Muttersprache definierte Sprache erstellt worden ist.

15

Vorteile und Zweckmäßigkeiten der Erfindung ergeben sich im übrigen aus der nachfolgenden Beschreibung eines Ausführungsbeispiels der Erfindung anhand der einzigen Figur. Diese zeigt ein schematisches Ablaufdiagramm der Eingabephase zur Erstellung eines Spracherkenner-Vokabulars gemäß der Erfindung.

20

Es soll die sprecherabhängige Namenswahl auf einem Mobiltelefon mit den Namen aus dem Telefonbuch für einen deutschsprachigen Benutzer realisiert werden. In dem Telefonbuch befinden sich neben überwiegend deutschsprachigen Namen auch einige fremdsprachige Namen. Ein Umsetzer für die graphemische Darstellung der Namen ist auf die Sprachen Deutsch, Italienisch, Tschechisch, Griechisch, Türkisch eingestellt, insgesamt als  $N = 5$  verschiedene Sprachen.

25

30

In einem Anfangsschritt S0 wird eine Sprachidentifikation der zugeführten Worte 10 bzw. Einträge des Telefonbuchs vorgenommen. Genauer gesagt wird jedes einzelne Wort auf die Wahrscheinlichkeit der Zugehörigkeit zu einer der fünf Sprachen analysiert. Wird beispielsweise ein deutscher Name verarbeitet, so wird die Wahrscheinlichkeit für Deutsch sehr hoch

35

sein, für die anderen vier Sprachen, nämlich Italienisch, Tschechisch, Griechisch und Türkisch dagegen sehr viel niedriger. Anhand der pro Wort ermittelten Wahrscheinlichkeiten wird die Sprache mit der geringsten Wahrscheinlichkeit für  
5 die weitere Verarbeitung weggelassen. Das bedeutet, daß im nachfolgenden Verarbeitungsgang nur noch vier, anstatt fünf Sprachen verarbeitet werden müssen.

10 In einem ersten Verfahrensschritt S1 wird für jedes Wort die phonetische Umschrift für jede der vier verschiedenen Sprachen ermittelt. Dadurch werden für jedes Wort vier Phonemfolgen entsprechend vier ersten Aussprachevarianten erhalten.

15 In einem zweiten Verfahrensschritt S2 wird anschließend eine Abbildung der Phoneme jeder der vier Sprachen auf den jeweiligen Phonemsatz der Muttersprache implementiert.

20 Diese Abbildung wird in einem dritten Verfahrensschritt S3 auf die im ersten Verfahrensschritt S1 erhaltenen vier ersten Phonemfolgen 12 angewandt. Hierdurch werden für jedes Wort vier zweite Phonemfolgen 14 entsprechend vier zweiten Aussprachevarianten erhalten. Die vier zweiten Phonemfolgen 14 können bereits mit einem Muttersprachen-Spracherkenner erkannt werden.

25

Um allerdings den Verarbeitungsaufwand für den Spracherkenner weiter zu reduzieren, wird pro Wort jede zweite Phonemfolge mittels der Levenshtein-Distanz analysiert und klassifiziert (Schritt S4). Anschließend folgt ein fünfter Verfahrensschritt S5, in dem die analysierten und klassifizierten zwei-  
30 ten Phonemfolgen pro Wort auf drei Phonemfolgen reduziert werden.

35 Schließlich wird in einem letzten Schritt S6 ein Spracherkenner-Vokabular mit dem in dem fünften Verfahrensschritt S5 erhaltenen drei zweiten Phonemfolgen pro Wort für den Muttersprachen-Spracherkenner erstellt. Durch die nochmalige Reduk-

tion der Phonemfolgen im fünften Verfahrensschritt S5 wird also das zu speichernde und während einer Spracherkennung zu durchsuchende Spracherkenner-Vokabular merklich reduziert. Dies bringt in einer praktischen Anwendung der Spracherken-  
5 nung einerseits den Vorteil eines geringeren Speicherplatzbedarfs und andererseits einer schnelleren Verarbeitung, da ein kleineres Vokabular durchsucht werden muß.

10 Nach Ablauf des beschriebenen Verfahrens kann der Benutzer mittels Spracherkennung eine Namenswahl, also den sprachgesteuerten Aufruf gespeicherter Rufnummern über den Namen des Teilnehmers vornehmen, ohne daß er den Namen des zu rufenden Teilnehmers explizit einmal vorsprechen, also trainieren, muß.

15 Im folgenden wird kurz erläutert, was der Benutzer des Mobiltelefons zur Verbesserung der Spracherkennung tun kann. Sollte er einmal feststellen, daß ein bestimmter Name nicht gut erkannt wird, kann er das Spracherkenner-Menü seines Mobiltelefons aufrufen und dort die Anwendung "Namenswahl" auswählen. Unter dieser Anwendung kann ihm nun eine oder auch mehrere Möglichkeiten angeboten werden, um die Spracherkennung eines bestimmten Wortes, genauer gesagt eines bestimmten Namens aus dem elektronischen Telefonbuch des Mobiltelefons zu  
20 verbessern. Im folgenden werden beispielhaft einige dieser Möglichkeiten kurz erläutert:

1. Der Benutzer kann das schlecht oder gar nicht erkannte Wort nochmals in das Mobiltelefon einsprechen und anschlie-  
30 ßend durch den im Mobiltelefon enthaltenen Spracherkenner in eine Phonemfolge umsetzen lassen. In diesem Fall werden vorher automatisch ermittelte Aussprachevarianten ganz oder teilweise, je nachdem welche Nähe sie zu der neu ermittelten Phonemfolge besitzen, aus dem Vokabular des Spracherkenners  
35 entfernt.

2. Alternativ kann sich der Benutzer auf dem Display des Mobiltelefons eine Art Lautschrift des schlecht oder gar nicht erkannten Eintrags des elektronischen Telefonbuchs anzeigen lassen. Die Art Lautschrift kann der Benutzer dann bei Nicht-  
5 zutreffen, d. h. bei schlechter Übereinstimmung mit seiner Aussprache, editieren. Beispielsweise kann durch die automatische Umsetzung des Eintrags "Jacques Chirac" als Lautschrift "Jakwes Schirack" gespeichert sein. Erscheint nun dem Benutzer diese Lautschrift als fehlerhaft, kann er sie mit-  
10 tels seines Mobiltelefons editieren, beispielsweise zu "Schack Schirack". Anschließend kann das System dazu die phonetische Beschreibung ermittelt und diese ins Spracherkenner-Vokabular neu aufnehmen. Damit sollte die automatische Spracherkennung zuverlässig funktionieren.
- 15 3. Schließlich kann der Benutzer durch eine explizite Angabe der Sprache, aus der ein fehlerhaft oder gar nicht erkannter Name stammt oder durch explizite Auswahl einer bestimmten Sprache für einen bestimmten Namen die Erkennung wesentlich  
20 verbessern. In einem derartigen Fall werden alle Aussprachevarianten für den Namen aus dem Spracherkenner-Vokabular entfernt, die nicht der explizit angegebenen Sprache zugeordnet sind.
- 25 Die Erfindung kann auch vorteilhaft in anderen mobilen Geräten außer einem Mobiltelefon z. B. einem Personal Assistant oder auch einem Personal Computer verwendet, d. h. eingesetzt werden.

EPO - Munich  
61

17. Jan. 2002

## Patentansprüche

1. Betriebsverfahren eines automatischen Spracherkenners zur sprecherunabhängigen Spracherkennung von Worten (10) aus verschiedenen Sprachen, insbesondere zur Erkennung von Namen aus verschiedenen Sprachen, das von einer als Muttersprache definierten Sprache ausgeht und eine Eingabephase zur Erstellung eines Spracherkenner-Vokabulars mit den folgenden Schritte aufweist:
- 10 (a) Ermitteln der phonetischen Umschriften von Wörtern jeweils für N verschiedene Sprachen, um pro Wort N erste Phonemfolgen (12) entsprechend N ersten Aussprachevarianten zu erhalten (S1),
- (b) Implementieren einer Abbildung der Phoneme jeder Sprache auf den jeweiligen Phonemsatz der Muttersprache (S2),
- 15 (c1) Anwenden der in Schritt (b) implementierten Abbildung auf die in Schritt (a) ermittelten N ersten Phonemfolgen (12) für jedes Wort, wodurch pro Wort N zweite Phonemfolgen (14) entsprechend N zweiten Aussprachevarianten erhalten werden,
- 20 die mit einem Muttersprachen-Spracherkenner erkannt werden können (S3), und
- (d) Erstellen eines Spracherkenner-Vokabulars mit den im vorhergehenden Schritt erhaltenen N zweiten Phonemfolgen pro Wort für den Muttersprachen-Spracherkenner (S6).
- 25
2. Betriebsverfahren nach Anspruch 1,
- ferner gekennzeichnet durch die folgenden, vor dem Schritt (d) und nach dem Schritt (c1) auszuführenden Schritte:
- (c2) Bearbeiten der N zweiten Phonemfolgen (14) entsprechend
- 30 den N zweiten Aussprachevarianten jedes Wortes, indem
- (c21) jede zweite Phonemfolge (14) mittels geeigneter Distanzen, insbesondere der Levenshtein-Distanz, analysiert und klassifiziert wird (S4), und
- (c22) die N zweiten Phonemfolgen jedes Wortes auf wenige,
- 35 vorzugsweise 2 bis 3, Phonemfolgen reduziert werden (S5).

3. Betriebsverfahren nach Anspruch 1 oder 2,  
dadurch gekennzeichnet, dass  
vor dem Schritt (a) eine Sprachidentifikation vorgenommen  
wird, durch die für jedes zu erkennende Wort die Wahr-  
5      scheinlichkeits der Zugehörigkeit zu jeder der N verschiedenen Spra-  
chen bestimmt wird, und anhand des Ergebnisses der Sprach-  
identifikation die Anzahl der in Schritt (a) zu verarbeiten-  
den Sprachen, vorzugsweise auf 2 bis 3 verschiedene Sprachen,  
reduziert werden, indem die Sprachen mit der geringsten Wahr-  
10      scheinlichkeit nicht weiter verarbeitet werden (S0).

4. Betriebsverfahren nach einem der Ansprüche 1 bis 3,  
dadurch gekennzeichnet, dass  
das Ermitteln der phonetischen Umschriften in Schritt (a)  
15      durch mindestens ein neuronales Netz erfolgt.

5. Betriebsverfahren nach einem der Ansprüche 1 bis 4,  
dadurch gekennzeichnet, dass  
als Muttersprachen-Spracherkenner ein Hidden-Markov-Modell  
20      eingesetzt wird, das für die als Muttersprache definierte  
Sprache erstellt worden ist.

6. Automatischer Spracherkenner zur sprecherunabhängigen  
Spracherkennung von Worten aus verschiedenen Sprachen, insbe-  
25      sondere zur Erkennung von Namen aus verschiedenen Sprachen,  
wobei eine der verschiedenen Sprachen als Muttersprache defi-  
niert ist, mit  
- einem Muttersprachen-Spracherkenner,  
- einem ersten Verarbeitungsmodul zum Ermitteln der phoneti-  
30      schen Umschriften von Wörtern jeweils für N verschiedene  
Sprachen, um pro Wort N erste Phonemfolgen entsprechend N  
ersten Aussprachevarianten zu erhalten,  
- einem zweiten Verarbeitungsmodul zum Implementieren einer  
Abbildung der Phoneme jeder Sprache auf den jeweiligen Pho-  
35      nemsatz der Muttersprache,  
- einem dritten Verarbeitungsmodul zum Anwenden der mit dem  
zweiten Verarbeitungsmodul implementierten Abbildung auf die

mit dem ersten Verarbeitungsmodul ermittelten N ersten Phonemfolgen für jedes Wort, wodurch pro Wort N zweite Phonemfolgen entsprechend N zweiten Aussprachevarianten erhalten werden, die mit dem Muttersprachen-Spracherkenner erkannt werden können, und

- einem vierten Verarbeitungsmodul zum Erstellen eines Spracherkenner-Vorkabulars mit den durch das dritte Verarbeitungsmodul erhaltenen N zweiten Phonemfolgen pro Wort für den Muttersprachen-Spracherkenner.

7. Automatischer Spracherkenner nach Anspruch 6, gekennzeichnet durch ein fünftes Verarbeitungsmodul zum Bearbeiten der N zweiten Phonemfolgen entsprechend den N zweiten Aussprachevarianten jedes Wortes, das derart ausgebildet ist, dass jede zweite Phonemfolge mittels geeigneter Distanzen, insbesondere der Levenshtein-Distanz, analysiert und klassifiziert wird, und die N zweiten Phonemfolgen jedes Wortes auf wenige, vorzugsweise 2 bis 3, Phonemfolgen reduziert werden.

8. Automatischer Spracherkenner nach Anspruch 6 oder 7, gekennzeichnet durch einen Sprachidentifikator, der vor das erste Verarbeitungsmodul geschaltet ist und für jedes zu erkennende Wort die Wahrscheinlichkeit der Zugehörigkeit zu jeder der N verschiedenen Sprachen bestimmt, und einen Sprachreduzierer, der die Anzahl der vom ersten Verarbeitungsmodul zu verarbeitenden Sprachen, vorzugsweise auf 2 bis 3 verschiedene Sprachen, reduziert werden, indem die Sprachen mit der geringsten Wahrscheinlichkeit nicht weiter verarbeitet werden.

9. Automatischer Spracherkenner nach einem der Ansprüche 6 bis 8,

dadurch gekennzeichnet, dass

das erste Verarbeitungsmodul mindestens ein neuronales Netz zum Ermitteln der phonetischen Umschriften aufweist.



10. Automatischer Spracherkenner nach einem der Ansprüche 6 bis 9,

dadurch gekennzeichnet, dass

5 der Muttersprachen-Spracherkenner ein Hidden-Markov-Modell aufweist, das für die als Muttersprache definierte Sprache erstellt worden ist.

10 11. Verwendung des Betriebsverfahrens nach einem der Ansprüche 1 bis 5 und des automatischen Spracherkenners nach einem der Ansprüche 6 bis 10 in einem mobilen Gerät wie einem Mobiltelefon, Personal Digital Assistant oder einem Personal Computer.

EPO - Munich  
61

17. Jan. 2002

## Zusammenfassung

Betriebsverfahren eines automatischen Spracherkenners zur  
sprecherunabhängigen Spracherkennung von Worten aus verschie-  
5 denen Sprachen und automatischer Spracherkenner

Die Erfindung betrifft ein Betriebsverfahren eines automati-  
schen Spracherkenners zur sprecherunabhängigen Spracherken-  
nung von Worten (10) aus verschiedenen Sprachen, insbesondere  
10 zur Erkennung von Namen aus verschiedenen Sprachen, das von  
einer als Muttersprache definierten Sprache ausgeht und eine  
Eingabephase zur Erstellung eines Spracherkenner-Vokabulars  
aufweist.

15 Figur

11/11

EPO - Munich  
61

17. Jan. 2002

